Dynamic time warping distance for message propagation classification in Twitter

Siwar Jendoubi[1,2,4], Arnaud Martin[2], Ludovic Liétard[2], Boutheina Ben Yaghlane[3], Hend Ben Hadji[4]

[1] LARODEC, ISG Tunis, Université de Tunis
[2] IRISA, Université de Rennes I
[3] LARODEC, IHEC Carthage, Université de Carthage
[4] Centre d'Etude et de Recherche des Télécommunications
`Siwar.Jendoubi@yahoo.fr`, `{Arnaud.Martin, ludovic.lietard}@univ-rennes1.fr`, `boutheina.yaghlane@ihec.rnu.tn`, `hend.Benhji@cert.mincom.tn`

Abstract. Social messages classification is a research domain that has attracted the attention of many researchers in these last years. Indeed, the social message is different from ordinary text because it has some special characteristics like its shortness. Then the development of new approaches for the processing of the social message is now essential to make its classification more efficient. In this paper, we are mainly interested in the classification of social messages based on their spreading on online social networks (OSN). We proposed a new distance metric based on the Dynamic Time Warping distance and we use it with the probabilistic and the evidential k Nearest Neighbors (k-NN) classifiers to classify propagation networks (PrNets) of messages. The propagation network is a directed acyclic graph (DAG) that is used to record propagation traces of the message, the traversed links and their types. We tested the proposed metric with the chosen k-NN classifiers on real world propagation traces that were collected from Twitter social network and we got good classification accuracies.

Keywords: Propagation network (PrNet), classification, Dynamic Time Warping (DTW), k Nearest Neighbor (k-NN).

1 Introduction

During the past decade, many classification methods have been appeared, like k Nearest Neighbors (k-NN), Naive Bayes, Support Vector Machines (SVM), etc. Those methods have been applied to several problems among them text classification and they proved their performance, [19]. However, when working with short text like online communications, chat messages, tweets, etc, we are face to a new challenge. In fact, in a short text there is no sufficient word occurrences or shared context for a good similarity measure. Let's take Twitter for example, Twitter is a micro-blogging service that allows its users to share messages of

140 characters that are called *tweets*. As a consequence, using a traditional text classification technique to classify tweets, like the "Bag-Of-Words" method, fail to achieve good classification rates due to the message shortness. Existing works on classification of short text integrate meta-information from external sources like Wikipedia, World Knowledge and MEDLINE [3, 11, 17]. They tend to enrich the content of the message.

The purpose of this paper is to classify social messages without any access to their content. Our work is motivated by two facts; first, it is not always possible to have access to the content of the message but we may have access to its propagation traces, in such a case, our approaches are useful. Another motivation is that, text processing techniques, always, need a pre-processing step in which it is necessary to remove URLs, stop words, questions, special characters, etc. When working with tweets, for example, after the pre-processing step, it falls, very often, on empty messages. Those empty messages can not be classified by a text based classification technique. Hence comes the necessity of new classification approaches that consider the propagation of the message.

Our work is driven by the motivations above, and it achieves the following contributions: 1) we adapted the Dynamic Time Warping (DTW) distance [16] to be used to measure the distance between two propagation networks (PrNet for short)[5]. 2) we proposed to incorporate the proposed distance in the probabilistic k-NN and the evidential k-NN [8] to classify propagation networks of social messages. Then 3) we tested the classifiers on real world propagation traces collected from Twitter social network.

This paper is organized as follow: Section 2 discusses some related works. Section 3 provides relevant background. Section 4 introduces the proposed PrNet-DTW distance. And in Section 5 presents results from our experiments.

2 Related works

2.1 Content based approaches

Methods that are used for text classification or clustering always have some limitation with short text, in fact, in short text there is no sufficient word occurrences. Then, traditional methods are not suitable for the classification of the social message that is characterized by its shortness. For example, the use of the traditional "Bag-Of-Words" method to classify tweets may fail to achieve good classification rates. This limitation has attracted the attention of many researchers who developed several approaches. The authors in [25] classified tweets to "News", "Events", "Opinions", "Deals" and "Private Messages" using a set of features among them author information and features extracted from the tweet. In [3] and [11], the authors propose approaches for short text clustering that use not only the content of the text but also an additional set of items that is extracted from an external source of information like Wikipedia and World

[5] We call propagation network the network that conserves propagation traces of the message, *i.e.* traversed links and nodes

Knowledge. Also, [17] classify short and sparse text using a large scale external data collected from Wikipedia and MEDLINE.

Social messages are, also, classified for sentiment analysis and opinion mining purposes [13]. The task here, is to identify the dominant opinion about a product or a brand using text mining techniques. The author of [14] used 3516 tweets to identify costumer's sentiment about some well known brands. In [10], authors used text published on Twitter and Facebook to analyze the opinion about three chain of pizza. The reader can refer to [15] for a recent survey.

Our work is different from all of the above in that we propose to classify the social message without access to its content. In fact, we predict the class of the message by interpreting its propagation traces through the social network. We think that the proposed approaches will be useful in the case where there is no access to the content of the message or when text based methods are unable to classify the message due to its shortness.

2.2 Propagation based approaches

Now we move to present two methods that were used to classify propagation networks and that were published in [12]. The first method uses the probability theory and the second one incorporates the theory of belief functions. As we said above, existing classification approaches that are used for text classification and characterization, always, have some limitation with short text. To overcome this limitation, we propose to classify the propagation traces of the message instead of its content. For an illustrative example, when you receive a letter from your bank, it is likely to be about your bank account.

The PrNet classifiers work in two main steps, the first step, is used to learn the model parameters and the second step, uses the learned model to classify new coming messages (propagation network of the message). Both methods have the same principle in the two steps. In the parameter learning step, we need a set of propagation networks, PrNetSet that is used to estimate a probability distribution defined on types of links for each level[6]. In the belief PrNet classifier, we use the consonant transformation algorithm, also called inverse pignistic transformation, [1, 2] that allows us to transform the probability distribution (output of the probabilistic parameter learning step) to a BBA distribution while preserving the least commitment principle [23]. Once model's parameters are learned, we can use it to classify a new message (propagation network of the message). The reader can refer to [12] for more details.

These classifiers need a transit step through a compact structure that assigns a probability distribution to each propagation level. This step leads to a loss of information that may be significant in the classification step. Another drawback is that these methods do not work with continuous types of links and a discretization step is always needed in such a case. We think that the proposed PrNet-DTW classifiers will avoid these problems.

[6] We call propagation level the number of links between the source of the message and the target node.

3 Background

3.1 Theory of belief functions

The *Upper and Lower probabilities* [7] is the first ancestor of the evidence theory, also called Dempster-Shafer theory or theory of belief functions. Then [20] introduced the *mathematical theory of evidence* and defined the basic mathematical framework of the evidence theory, often called *Shafer model*. The main goal of the Dempster-Shafer theory is to achieve more precise, reliable and coherent information.

Let $\Omega = \{s_1, s_2, ..., s_n\}$ be the frame of discernment. The basic belief assignment (BBA), m^{Ω}, represents the agent belief on Ω. $m^{\Omega}(A)$ is the mass value assigned to $A \subseteq \Omega$, it must respect: $\sum_{A\subseteq\Omega} m^{\Omega}(A) = 1$. In the case where we have $m^{\Omega}(A) > 0$, A is called focal set of m^{Ω}.

Combination rules are the main tools that can be used for information fusion. In fact, in real world applications, we do not have the same kind of information to be combined, that's why the same combination rule may performs well in some applications and may gives unsatisfiable results with other applications. Among these combination rules, we find the Dempster's rule [7], the conjunctive rule of combination (CRC) [21, 22] and the disjunctive rule of combination (DRC) [22].

3.2 *k* Nearest Neighbors

In this paper, we choose the k nearest neighbors classification technique because it is distance based. It will be used to classify propagation traces of social messages together with the proposed distance. In this section we present two k-NN based approaches which are the probabilistic k-NN and the evidential k-NN.

Probabilistic k nearest neighbors (k-NN) is a well known supervised method that is generally used for classification. It needs as input a set of training examples that we know their features values and their classes, and of course the object to be classified. Besides we have to specify a measure of distance that will be used to quantify the matching between the new object x and every object in the training set. First, the k-NN starts by computing the distance between x and every object in the training set, then, it selects the k nearest neighbors, *i.e.* that have the shortest distance with x. Finally, the object x is classified according to the majority vote principle, *i.e.* the algorithm chooses the class that has the maximum occurrence count in the k nearest neighbors set to be the class of x. The k-NN technique is surveyed in [5].

Evidential k Nearest Neighbors is an extension of the probabilistic k-NN to the theory of belief functions [8]. The probabilistic k-NN uses distances between the object x, to be classified, and objects in the training set to sort the training example, then it chooses the k nearest neighbors to x. However, according to [8], the distance value between x and its nearest neighbors may be significant. The evidential k-NN differs from the probabilistic one in the decision rule. Let $\Omega = \{s_1, s_2, ..., s_n\}$ the set of all possible classes, be our frame of discernment and d_j be the distance between x and the j^{th} nearest neighbor. The idea behind the evidential k-NN consists on representing each object of the k neighbors by a BBA distribution defined by:

$$m\left(\{s_i\}\right) = \alpha \tag{1}$$
$$m\left(\Omega\right) = 1 - \alpha \tag{2}$$
$$m\left(A\right) = 0\,\forall A \in 2^C \setminus \{C_i\} \tag{3}$$

such that $0 < \alpha < 1$. If d_j is big, α have to be small. Then it will be calculated as follow:

$$\alpha = \alpha_0 \Phi_i\left(d_j\right) \tag{4}$$
$$\Phi_i\left(d_j\right) = e^{-\gamma_i d_j^{\beta}} \tag{5}$$

where $\gamma_i > 0$ and $\beta \in \{1, 2, \ldots\}$. After estimating a BBA distribution for each nearest neighbor, the decision about the class of x is made according to the following steps; first we combine all BBA distributions using a combination rule. Second, we apply the pignistic transformation, [24], in order to obtain a pignistic probability distribution. And finally, we choose the class that have the biggest pignistic probability. In the next section, we will introduce the dynamic time warping distance and its extension to compute similarity between propagation networks.

4 Proposed dynamic time warping distance for propagation networks similarity

The propagation network is a graph based data structure that is used to store propagation traces of a message. The PrNet has two main characteristics that distinguish it from an ordinary DAG[7]; first, its arcs are weighted by the type of the relationship between users, and second, its paths are time dependent. In this paper, we choose to use distance based classifiers; the probabilistic and the evidential k-NN, then, we need to measure the distance between the PrNet to be classified and the training set. In [12], we presented two PrNet classifiers that are based on mathematical distances like the Euclidean distance and the Jaccard distance. This solution need to transform the PrNet to a set of probability or BBA distributions, then it computes the distance between those distributions instead of PrNets. This transformation may lead to a loss of the information. A second solution may be to use a graph distance metric to measure the similarity between PrNets. In the literature, we found several distances like *Graph edit distances* [9], and *Maximal common sub-graph based distances* [6]. However, all these distances do not consider the time dimension which is a character of the PrNet. Then comes the need of a new distance that is adapted to weighted time dependent DAGs like the PrNet. As a solution to this problem we propose the Dynamic Time Warping distance for propagation networks similarity (PrNet-DTW).

[7] Directed Acyclic Graph

The Dynamic Time Warping similarity measure [18] was first proposed for speech recognition, it consider the fact that the speech is time dependent. Recently, [16] propose to use it to measure the similarity between two sequences, *i.e.* a sequence is an ordered list of elements. DTW distance is used to consider the order of appearance of each element in the sequences while computing the distance between them. Let $A = (a_1, a_2, \ldots, a_S)$ and $B = (b_1, b_2, \ldots, b_T)$ be two sequences. $DTW\,(A_i, B_j)$ is the DTW distance between A and B and it is defined as [16]:

$$DTW\,(A_i, B_j) = \delta\,(a_i, b_j) + \min \begin{cases} DTW\,(A_{i-1}, B_{j-1}) \\ DTW\,(A_i, B_{j-1}) \\ DTW\,(A_{i-1}, B_j) \end{cases} \tag{6}$$

Note that $\delta\,(a_i, b_j)$ is a the distance between the two elements $a_i \in A$ and $b_j \in B$. As mentioned in [16], the implementation of this recursive function leads to exponential temporal complexity. They propose the memoization technique as a solution to speed up the computation. Hence, we need a $\mid S \mid \times \mid T \mid$ matrix in which we record previous results in order to avoid their computation in next iterations. This computation technique maintain the time and space complexity of the DTW distance to $O\,(\mid S \mid \times \mid T \mid)$.

The PrNet-DTW distance is used to measure the distance between two propagation networks. In the first step, we transform each PrNet to a set of dipaths. We define a dipath as a finite sequence vertices connected with arcs that are directed to the same direction (line 1 and 2 in algorithm 1). We note that all dipaths starts from the source of the message. In the second step, the PrNet-DTW algorithm loops on the DipathSet1, at each iteration, it fixes a Dipath and compute its DTW distance with all Dipaths in DipathSet2 and it takes the minimal value. Finally, it computes the mean of minimal distances between Dipaths in DipathSet1 and those in DipathSet2 to be the PrNet-DTW distance. Details are shown in algorithm 1. We choose the k-NN algorithm and evidential k-NN algorithm to classify propagation networks because they are distance based classifiers and they can be used with the proposed PrNet-DTW distance.

Algorithm 1: PrNet-DTW algorithm

input : *PrNet1* and *PrNet2*: Two propagation networks
output: *Distance:* The distance between PrNet1 and PrNet2.
begin
1 $DipathSet1 \leftarrow PrNet1.TransformToDipathSet()$
2 $DipathSet2 \leftarrow PrNet2.TransformToDipathSet()$
3 **for** $i = 1$ **to** $DipathSet1.size()$ **do**
4 $D \leftarrow maxValue$
5 **for** $j = 1$ **to** $DipathSet2.size()$ **do**
6 $D \leftarrow \min\,(D,\ DTW\,(DipathSet1.get(i),\ DipathSet2.get(j)))$
7 $Distance \leftarrow Distance + D$
8 $Distance \leftarrow Distance/DipathSet1.Size\,();$

5 Experiments and results

We used the library Twitter4j[8] which is a java implementation of the Twitter API to collect Twitter data. We crawled the Twitter network for the period between 08/09/2014 and 03/11/2014. After a data cleaning step, we got our data set that contains tweets of three different classes: "Android", "Galaxy" and "Windows". To simplify the tweet classification step, we consider a tweet that contains the name of a class C, for example a tweet that contains the word "Android", of type that class C, i.e. the class "Android" in our example. Table 1 presents some statistics about the data set.

Table 1: Statistics of the data set

	#User	**#Follow**	**#Tweet**	**#Retweet**	**#Mention**	**#Prop. links**	**#PrNet**
Android	6435	9059	81840	3606	6092	7623	224
Galaxy	4343	4482	8067	2873	5965	6819	161
Windows	5775	12466	11163	2632	3441	11400	219

The remainder of this section is organized as follow: we present our experiments configuration, the method with which we extracted propagation and the computation process of link weights. Then, we compare the proposed classifiers with those of [12].

5.1 Experiments configuration

In our experiments, we need to extract propagation traces of each type of message. Here, we consider that a tweet of type a was propagated from a user u to a user v if and only if u posts a tweet of type a before v and at least one of these relations between u and v exists: 1) v follows u, 2) u mentions v in a tweet of type a, 3) v retweets a tweet of type a written by u. After getting propagation traces we extract propagation networks such that each PrNet has to have one source.

We define types of links that are used to measure the similarity between propagation networks. In Twitter social network there are three possible relations the first one is explicit which is the follow relation, the second and the third relations are implicit which are the mention and the retweet. Another property of Twitter, is that between two users u and v we can have a follow, a mention and/or a retweet relation. We assign to each of those a weight [4] and we assign to each link a vector of weights that has the form (w_f, w_m, w_r). Let S_u be the set of successor of u, P_u the set of predecessor of u, T_u the set of tweets of u, $R_u(v)$ the set of tweets of u that were retweeted by v, $M_u(v)$ the set of tweets of u in which v was mentioned and M_u the set of tweets in which u mentions another user. We compute weights [4] as follow:

[8] Twitter4j is a java library for the Twitter API, it is an open-sourced software and free of charge and it was created by Yusuke Yamamoto. More details can be found in http://twitter4j.org/en/index.html.

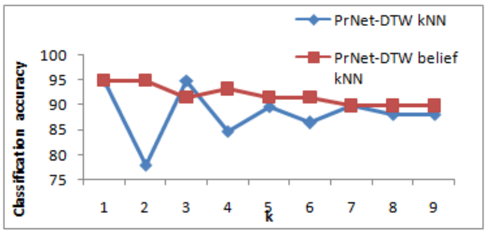

Fig. 1: k variation

- Follow relation: $w_f\left(u,v\right) = \frac{|S_u \cap (P_u \cap \{u\})|}{|S_u|}$
- Mention relation: $w_m\left(u,v\right) = \frac{|M_u(v)|}{|M_u|}$
- Retweet relation: $w_r\left(u,v\right) = \frac{|R_u(v)|}{|T_u|}$

Finally, we choose the euclidean distance to evaluate the $\delta\left(a_i, b_j\right)$ in the computation process of the PrNet-DTW.

5.2 Experiments evaluation

In our experiments, we want to evaluate the performance of the PrNet-DTW distance, then, we integrate it in the k-NN and the evidential k-NN classifiers and we compare the proposed classifiers with those proposed in [12]. As PrNet classifiers works with a discrete types of links [12], a discretization step was needed, *i.e.* if the weight value (w_f, w_m or w_r) is greater than 0 we replace it by 1 in the discrete weight vector elsewere we replace it by 0. For example, if the link is weighted by the vector ($w_f = 0.5$, $w_m = 0$, $w_r = 0.25$), the output after the discretization step will be (1, 0, 1). In the remainder of our experiments, we divide, randomly, our data set into two subsets; the first one contains 90% of PrNets and it is used for training and the second one (10%) is used for testing.

The algorithm k-NN is known to be dependent to k value, and varying k may vary the classification accuracy. Then, to see the impact of the parameter k, we made this experiment; we run our k-NN based algorithms with multiple k values and we obtained results in Figure 1. We note that odd values are more appropriate to k when we use PrNet-DTW Probabilistic k-NN. Moreover, the PrNet-DTW belief k-NN has not the same behavior as the PN-DTW Probabilistic k-NN. In fact, the curve of the evidential classifier is more stable than the curve of the probabilistic one and the variation of the value of k does not have a great effect on the classification accuracy.

A second experiment was done to evaluate and compare the proposed classification methods. We fixed the parameter k to 5 and we obtained results in table 2. As shown in table 2, the probabilistic and the belief classifiers do not give good classification accuracy, this behavior is a consequence of the discretization step that leads to the loss of the information given by weights values. In contrast, the PrNet-DTW based classifiers show their performance, indeed, we have got good accuracy rates: 88.69% (±3.39, for a 95% confidence interval) and

Table 2: Comparison between PrNet classifiers

	Proba classifier	**Belief classifier**	**PrNet-DTW** k**-NN**	**PrNet-DTW Belief** k**-NN**
Accuracy	51.97% ±2.04	52.25% ±1.99	**88.69%** ±3.39	**89.92%** ±3.20

89.92% (±3.20) respectively. We see also that the PrNet-DTW belief classifier gives slightly better results.

6 Conclusion

To sum up, we presented a new distance metric that we called PrNet-DTW. Our measure is used to quantify the distance between propagation networks. Also, we showed the performance of our measure in the process of classification of propagation networks, indeed, we defined two classification approaches that uses the PrNet-DTW measure which are the probabilistic k-NN and the evidential k-NN.

For future works, we will search to improve the PrNet-DTW based classifiers by taking into account the content of the message to be classified, in fact, we believe that a classification approach that uses information about the content of the message and information about its propagation will further improve the results.

7 Acknowledgement

These research works and innovation are carried out within the framework of the device *MOBIDOC* financed by the European Union under the *PASRI* program and administrated by the *ANPR*. Also, we thank the "*Centre d'Etude et de Recherche des Télécommunications*" (CERT) for their support.

References

1. Aregui, A., Denœux, T.: Fusion of one-class classifiers in the belief function framework. In: Proc. of FUSION. Québec, Canada (juillet 2007)
2. Aregui, A., Denoeux, T.: Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities. Int. J. of Approximate Reasoning 49(3), 575–594 (2008)
3. Banerjee, S., Ramanathan, K., Gupta, A.: Clustering short texts using wikipedia. In: Proc. of ACM SIGIR Conf. pp. 787–788. ACM (2007)
4. Ben Jabeur, L.: Leveraging social relevance: Using social networks to enhance literature access and microblog search. Ph.D. thesis, Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier) (October 2013)
5. Bhatia, N., Vandana: Survey of nearest neighbor techniques. IJCSIS 8(2), 302–305 (2010)

6. Bunke, H., Foggia, P., Guidobaldi, C., Sansone, C., Vento, M.: A comparison of algorithms for maximum common subgraph on randomly connected graphs. In: International Workshop SSSPR. pp. 123–132. Springer (August 2002)
7. Dempster, A.P.: Upper and Lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
8. Denœux, T.: A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 25(5), 804–813 (mai 1995)
9. Gao, X., Xiao, B., Tao, D., Li, X.: A survey of graph edit distance. Int. J. of Future Computer and Communication 13(1), 113–129 (Février 2010)
10. He, W., Zhab, S., Li, L.: Social media competitive analysis and text mining: A case study in the pizza industry. Int. J. of Information Management 33, 464–472 (2013)
11. Hu, X., Sun, N., Zhang, C., Chua, T.S.: Exploiting internal and external semantics for the clustering of short texts using world knowledge. In: Proc. of CIKM. pp. 919–928. ACM (2009)
12. Jendoubi, S., Martin, A., Lietard, L., Yaghlane, B.B.: Classification of message spreading in a heterogeneous social network. In: Proc. of IPMU 2014 (July 2014)
13. Lo, Y.W., Potdar, V.: A review of opinion mining and sentiment classification framework in social networks. In: Proc. of DEST'09 (Juin 2009)
14. Mostafa, M.M.: More than words: Social networks text mining for consumer brand sentiments. Expert Systems with Applications 40, 4241–4251 (2013)
15. Othman, M., Hassan, H., Moawad, R., El-Korany, A.: Opinion mining and sentimental analysis approaches: A survey. Life Science Journal 11(4), 321–326 (2014)
16. Petitjean, F., Inglada, J., Gancarski, P.: Satellite image time series analysis under time warping. IEEE Transactions on Geoscience and Remote Sensing 50(8), 3081–3095 (2012)
17. Phan, X.H., Nguyen, L.M., Horiguchi, S.: Learning to classify short and sparse text and web with hidden topics from large-scale data collections. In: Proc. of WWW'09. pp. 91–100. ACM (2009)
18. Sakoe, H., Chiba, S.: A dynamic programming approach to continuous speech recognition. Proc. of the Seventh International Congress on Acoustics, Budapest 3, 65–69 (1971)
19. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47 (2002)
20. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
21. Smets, P.: The Combination of Evidence in the Transferable Belief Model. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(5), 447–458 (1990)
22. Smets, P.: Belief Functions: the Disjunctive Rule of Combination and the Generalized Bayesian Theorem. Int. J. of Approximate Reasoning 9, 1–35 (1993)
23. Smets, P.: Data Fusion in the Tranferable Belief Model. In: Proc. of FUSION. vol. 1, pp. 21–33. Paris, France (juillet 2000)
24. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. Int. J. of Approximate Reasonning 38, 133–147 (2005)
25. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proc. of ACM SIGIR. pp. 841–842. ACM (2010)